

Sections 9.2, 9.3

Law of large numbers (LLN),

Central limit theorem (CLT):

General versions

Notation: We say r.v.s $X_1, X_2, \dots, X_n$ are *independent and identically distributed* (iid) if they each have the same distribution and are independent.

Ex) If $X \sim \text{Bin}(n, p)$, $X = X_1 + \dots + X_n$ where $X_1, \dots, X_n$ are iid,
and $X_i \sim \text{Bern}(p)$. shorthand: $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$.

Basic message today: LLN and CLT hold for almost any sum of iid r.v.s.

Law of large numbers (LLN)

Theorem (LLN): Let $X_1, X_2, \dots, X_n$ be iid r.v.s with $\mathbb{E}X_i = \mu$, $\text{Var}(X_i) = \sigma^2 < \infty$. Let

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

be the sample average. Then for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \frac{P(|\bar{X}_n - \mu| \geq \epsilon)}{1} = 0 \quad \text{or, equivalently,} \quad \lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$$

Meaning:

- For n large, $\bar{X}_n$ is essentially constant (the constant is μ).
- Sample average $\rightarrow$ Expected value

Ex) You roll n dice and let $\bar{X}_n$ be the average. Then, for n large enough

$$P(|\bar{X}_n - 3.5| < \frac{10^{-10}}{n}) \geq 99.999999\%$$

Also,

$$\begin{aligned} \lim_{n \rightarrow \infty} P(\bar{X}_n \geq 3.51) &= \frac{P(\bar{X}_n - 3.5 \geq 0.01)}{n} \\ &\leq \frac{P(|\bar{X}_n - 3.5| \geq 0.01)}{n} \\ &\xrightarrow{n \rightarrow \infty} 0 \quad \text{by LLN} \end{aligned}$$

Proof of LLN

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$$

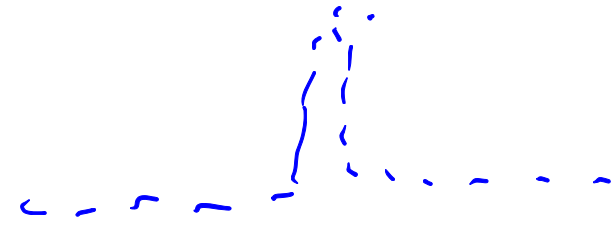
Recall: $\mathbb{E}X_i = \mu$, $\text{Var}(X_i) = \sigma^2$ so $\mathbb{E}\bar{X}_n = \underline{\mu}$, $\text{Var}(\bar{X}_n) = \underline{\frac{\sigma^2}{n}}$

As $n \rightarrow \infty$, we wish to show

$$P(|\bar{X}_n - \mu| \geq \epsilon) \rightarrow 0.$$

Indeed, by Chebyshev inequality

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n \rightarrow \infty} 0 \quad \checkmark$$



LLN: $\bar{X}_n \rightarrow \mu$ as $n \rightarrow \infty$ (it converges in probability)

This occurs because $\text{Var}(\bar{X}_n) \sim \frac{1}{n} \rightarrow 0$.

Q) If you focus on a small interval around μ , what does the distribution of $\bar{X}_n$ look like?

CLT will examine $Z_n = \frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu) = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu)$.

Note: $\text{Var}(Z_n) = 1$, ~~$\neq Z_n = 0$~~ .

Central limit theorem (CLT)

Theorem (CLT): Let $X_1, X_2, \dots, X_n$ be iid r.v.s with $\mathbb{E}X_i = \mu$, $\text{Var}(X_i) = \sigma^2 < \infty$. Let

$$Z_n = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu).$$

Then $Z_n \xrightarrow{\text{dist}} N(0,1)$ i.e., for any $-\infty \leq a \leq b \leq \infty$,

$$\lim_{n \rightarrow \infty} \underbrace{P(Z_n \in [a, b])}_{\text{green wavy line}} = P(N(0,1) \in [a, b])$$

can be open interval, or half open

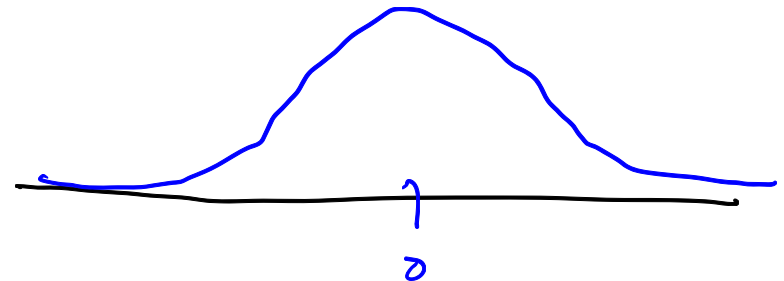
Interpreting CLT

- $Z_n = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n (X_i - \mu)$ satisfies $\mathbb{E}Z_n = \underline{\quad 0 \quad}$

$\text{Var}(Z_n) = \underline{\quad 1 \quad}$

- Meaning: A normalized sum of many iid r.v.s is almost $N(0,1)$.

$$\frac{1}{\sqrt{n}} \sum \pm 1 \approx$$



Normal approximation (general)

Let $X_1, X_2, \dots, X_n$ be iid with $\mathbb{E}X_i = \mu$, $\text{Var}(X_i) = \sigma^2$.

Let $S_n = X_1 + \dots + X_n$, so $\mathbb{E}S_n = \underline{n\mu}$, $\text{Var}(S_n) = \underline{n\sigma^2}$.

Then, if n is large, $S_n \approx N(\underbrace{\underline{n\mu}}_{\mathbb{E}S_n}, \underbrace{\underline{n\sigma^2}}_{\text{Var}(S_n)})$

Ex) Roll 2000 dice. Approximate the probability that the sum is at least 8000.

$$S_{2000} = X_1 + \dots + X_{2000}$$

$$\begin{aligned} \mathbb{E} S_{2000} &= 2000 \cdot \mathbb{E} X_1 \\ &= 2000 \cdot \frac{7}{2} = 7000 \end{aligned}$$

$$\begin{aligned} \text{Var}(S_{2000}) &= 2000 \text{Var}(X_1) \\ &= 2000 \left(\frac{91}{6} - \frac{49}{4} \right) \end{aligned}$$

Normal approx: $S_{2000} \approx N\left(7000, 2000 \left(\frac{91}{6} - \frac{49}{4}\right)\right)$

$$P(S_{2000} \geq 8000) \approx P\left(N\left(7000, 2000 \left(\frac{91}{6} - \frac{49}{4}\right)\right) \geq 8000\right)$$

$$= P\left(7000 + \sqrt{2000 \left(\frac{91}{6} - \frac{49}{4}\right)} N(0,1) \geq 8000\right)$$

$$= P\left(N(0,1) \geq \frac{1000}{\sqrt{2000 \left(\frac{91}{6} - \frac{49}{4}\right)}}\right) = \left[1 - \Phi\left(\frac{1000}{\sqrt{2000 \left(\frac{91}{6} - \frac{49}{4}\right)}}\right)\right]$$

Further topics (not tested)

- Error bound in CLT
- Practical ramifications
- Tail bounds for $N(0,1)$
- Proof sketch of CLT

Error bound in CLT

Berry-Esseen CLT

$$S_n = X_1 + \dots + X_n$$

$$X_i \stackrel{\text{dist}}{=} X$$

Theorem: $|P(Z_n \leq x) - \Phi(x)| \leq \frac{3\mathbb{E}|X-\mu|^3}{\sigma^3\sqrt{n}} \sim \frac{1}{\sqrt{n}}$

$$\text{Ex) } S_n = X_1 + \dots + X_n$$

$$X_i = \begin{cases} 1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases}$$

$$\mathbb{E}[X_i] = 0, \quad \text{Var}(X_i) = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2$$

$$= 1$$

$$\mathbb{E}[X_i - \mu]^3 = \mathbb{E}[X_i]^3 = 0$$

$$\text{RHS} = \frac{3 \cdot 1}{1\sqrt{n}} = \frac{3}{\sqrt{n}} \Rightarrow \left| P\left(\frac{\sum_{i=1}^n X_i}{\sqrt{n}} \leq t\right) - \Phi(t) \right| \leq \frac{3}{\sqrt{n}}, \quad t \in \mathbb{R}$$

Practical ramifications

- Normal distribution is used as a standard model in practice (even when an underlying sum of iid r.v.s is not identified).
 - In particular, it is the standard model for noise in the measurements/model
- Normal distribution has very quickly decaying tails, and thus, so do sums of iid r.v.s. Recall:

Fact: A $N(0,1)$ r.v. is very unlikely to deviate far from 0.

$$P(-1 \leq N(0,1) \leq 1) \approx 0.68$$

$$P(-2 \leq N(0,1) \leq 2) \approx 0.95$$

$$P(-3 \leq N(0,1) \leq 3) \approx 0.997$$

Tail bound for normal r.v.

Proposition: $P(N(0,1) \geq t) \leq e^{-\frac{t^2}{2}}$

Proof: Let $Z \sim N(0,1)$. Consider mgf

$$P(Z \geq t) = P(e^{\lambda Z} \geq e^{\lambda t})$$

$\lambda \in \mathbb{R}$ to be tuned

$$\leq \frac{\mathbb{E} e^{\lambda Z}}{e^{\lambda t}} = \frac{M_Z(\lambda)}{e^{\lambda t}}$$

by Markov $\neq$

$$= e^{\lambda^2/2} \cdot e^{-\lambda t} = \exp(\lambda^2/2 - \lambda t) = (*)$$

$$\frac{d}{d\lambda} \left(\frac{\lambda^2}{2} - \lambda t \right) = \lambda - t = 0 \Rightarrow \lambda = t \text{ is optimal. Plug in a.b.v.t.:$$

$$(*) = \exp(t^2/2 - t^2) = \exp(-t^2/2) \quad \checkmark$$

Proof sketch of CLT

For simplicity, assume that $\mathbb{E}X_i = 0$, $\text{Var}(X_i) = 1$. Recall, $Z_n = \frac{1}{\sqrt{n}} \sum_i X_i$.

$$\text{MGF of } Z_n: M(t) = \mathbb{E}\left[\exp\left(t Z_n\right)\right] = \mathbb{E}\left[\exp\left(\frac{t}{\sqrt{n}} \sum_{i=1}^n X_i\right)\right]$$

$$= \mathbb{E}\left[\underbrace{e^{\frac{t}{\sqrt{n}} X_1}}_{\text{indep}} \cdot \underbrace{e^{\frac{t}{\sqrt{n}} X_2}}_{\text{indep}} \cdots \underbrace{e^{\frac{t}{\sqrt{n}} X_n}}_{\text{indep}}\right]$$

$$= \mathbb{E}\left[e^{\frac{t}{\sqrt{n}} X_1}\right] \cdot \mathbb{E}\left[e^{\frac{t}{\sqrt{n}} X_2}\right] \cdots \mathbb{E}\left[e^{\frac{t}{\sqrt{n}} X_n}\right]$$

$$= \left(\mathbb{E}\left[e^{\frac{t}{\sqrt{n}} X_1}\right]\right)^n \quad \text{since } X_1 \stackrel{\text{dist}}{=} X_2 \stackrel{\text{dist}}{=} \cdots \stackrel{\text{dist}}{=} X_n$$

$$= \left(\mathbb{E} \left[1 + \frac{t}{\sqrt{n}} X_1 + \frac{\left(\frac{t}{\sqrt{n}}\right)^2}{2} X_1^2 + \mathcal{O}\left(\frac{1}{n^{3/2}}\right) \right] \right)^n$$

Taylor expansion

$$= \left[1 + \frac{t}{\sqrt{n}} \cancel{\mathbb{E} X_1} + \frac{t^2}{2n} \cancel{\mathbb{E} X_1^2} + \underbrace{\mathcal{O}\left(\frac{1}{n^{3/2}}\right)}_{\text{insig as } n \rightarrow \infty} \right]^n$$

$$\approx \left(1 + \frac{t^2}{2n} \right)^n \xrightarrow{n \rightarrow \infty} \underbrace{e^{t^2/2}}_{\text{MGF of } N(0,1)}$$

Convergence of MGF $\Rightarrow$ conv. of dist.!